

Quartic scaling MP2 for solids: A highly parallelized algorithm in the plane-wave basis

Tobias Schäfer, Benjamin Ramberger, and Georg Kresse
University of Vienna, Faculty of Physics and Center for Computational
Materials Science, Sensengasse 8/12, A-1090 Vienna, Austria

We present a low-complexity algorithm to calculate the correlation energy of periodic systems in second-order Møller-Plesset perturbation theory (MP2). In contrast to previous approximation-free MP2 codes, our implementation possesses a quartic scaling, $\mathcal{O}(N^4)$, with respect to the system size N and offers an almost ideal parallelization efficiency. The general issue that the correlation energy converges slowly with the number of basis functions is solved by an internal basis set extrapolation. The key concept to reduce the scaling of the algorithm is to eliminate all summations over virtual bands which can be elegantly achieved in the Laplace transformed MP2 (LTMP2) formulation using plane-wave basis sets. Analogously, this approach could allow to calculate second order screened exchange (SOSEX) as well as particle-hole ladder diagrams with a similar low complexity. Hence, the presented method can be considered as a step towards systematically improved correlation energies.

I. INTRODUCTION

When calculating the ground state energy of matter in a perturbative approach, the second-order Møller-Plesset perturbation theory (MP2) [1] is the lowest order correction to the well-established Hartree-Fock (HF) approximation. This correction includes electron correlation and non-covalent effects like van der Waals interaction, making MP2 very attractive for ab initio calculations of molecules and nonmetallic solids. However, in the traditional MP2 formulation [2], the improvements compared to HF come along with a very high price of computational effort. The scaling of the computation time with respect to the number of basis functions is quintic, $\mathcal{O}(N^5)$, if no further approximation is used.

To overcome this steep scaling, several attempts have been made for finite systems like molecules [3–7], where linear scaling codes, $\mathcal{O}(N)$, could be achieved using approximations based on the locality of one-electron orbitals, local MP2 (LMP2), and Gaussian basis sets. Nonetheless, it is generally agreed that these methods can suffer from accuracy problems [8] and slow down considerably when improved diffuse basis sets are used.

For periodic systems like crystalline solids, growing interest in MP2 calculations can be noticed [9–27]. Like for molecules there are plenty of MP2 implementations available for periodic systems today. While applications for specific extended 1D and 2D systems go back to Suhai [28], and Sun and Bartlett [29], the first general purpose MP2 computer program for periodic systems was CRYSCOR by Pisani et al. [30]. Based on the LMP2 approach for molecules, CRYSCOR inherits the linear scaling, $\mathcal{O}(N)$, but also the mentioned accuracy and basis set issues. The first general purpose MP2 implementation for periodic systems without such accuracy issues was implemented in VASP [31, 32] by Marsman, Grüneis et al. [33, 34] using a plane-wave basis set. Yet, this implementation sustains the unfavorable quintic scaling, $\mathcal{O}(N^5)$, making it difficult to treat large systems with more than 64 atoms, both regarding memory and computation time.

Further methods for periodic systems balancing computational effort and accuracy rely on the resolution-of-identity approximation (RI) [35], both RI and localized atomic orbitals [36], or real-space Monte Carlo integration of a Green’s function MP2 formulation [37, 38]. Finally, a high performance code for finite and periodic MP2 calculations became available quite recently, providing a high parallelization efficiency. Implemented in CP2K by VandeVondele and co-workers [39], this high performance approach possesses a reduced prefactor for a, still, largely quintic scaling, $\mathcal{O}(N^5)$.

Here we present a novel implementation of a high performance algorithm for exact MP2 calculations of periodic systems that provides a very high parallelization efficiency with low memory requirements and the computation time scales only quartic with the system size, $\mathcal{O}(N^4)$. The lower scaling is achieved by the Laplace-transformed MP2 [40] formulation (LTMP2) allowing for a presummation over all virtual orbitals. The code is implemented in VASP using plane-wave basis sets. The high parallelization efficiency is attained by dividing the set of all plane-waves over all CPUs, leading to a communication-free distribution of an outer plane-wave loop.

In this paper we begin with a theoretical part, Sec. II, containing a brief introduction to the canonical MP2 formulation, a schematic summary of the main strategy to reduce the scaling, and a comprehensive derivation of the quartic scaling LTMP2 formulation for periodic systems. In Sec. III the implementation is described regarding the parallelization strategy and the internal basis set extrapolation. We also provide pseudocode for the serial and parallel version. Benchmark calculations can be found in Sec. IV. We show the measured scaling behavior in both computation time and memory, the parallelization efficiency, as well as the numerical agreement of the resulting MP2 energies of our new code and the previous MP2 code of VASP. Lithium hydride (LiH) and methane (CH_4) in a chabazite cage serve as benchmark systems.

II. THEORY

In Møller-Plesset perturbation theory [1] the correlation energy E_c is estimated by means of a Rayleigh-Schrödinger perturbation series of the ground state energy

$$E = E^{(0)} + E^{(1)} + E^{(2)} + \dots \quad (1)$$

The correlation energy is conventionally defined as the difference between the ground state energy and the Hartree-Fock energy, since the latter neglects correlation effects: $E_c = E - E_{\text{HF}}$. Starting from a Hartree-Fock Hamiltonian, H_{HF} , the full Hamiltonian, H , is considered as the Hartree-Fock Hamiltonian complemented by a perturbation:

$$H = H_{\text{HF}} + H_c, \quad H_c = H - H_{\text{HF}}. \quad (2)$$

The perturbation H_c contains the full electron-electron interaction minus the electron-mean field interaction. Applying the common formulae of Rayleigh-Schrödinger perturbation theory, the sum of the zeroth and first order term of the perturbation series turns out to be the Hartree-Fock energy: $E_{\text{HF}} = E^{(0)} + E^{(1)}$. Thus the leading contribution to the correlation energy is given by the second order term (MP2 energy): $E_c = E^{(2)} + \dots$. The textbook formula [2] reads:

$$E^{(2)} = \frac{1}{2} \sum_{ij}^{\text{occ.}} \sum_{ab}^{\text{virt.}} \frac{\langle ij|ab \rangle [\langle ab|ij \rangle - \langle ab|ji \rangle]}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}. \quad (3)$$

The orbitals $|i\rangle / |a\rangle$ and energies $\varepsilon_i / \varepsilon_a$ are the occupied / virtual solutions of the Hartree-Fock equations respectively. With $\langle ab|ij \rangle$ we identify a two-electron integral,

$$\langle ij|ab \rangle = \int d^3r \int d^3r' \frac{\varphi_i^*(\mathbf{r}) \varphi_j^*(\mathbf{r}') \varphi_a(\mathbf{r}) \varphi_b(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad (4)$$

which is simply a matrix element of the two-electron Coulomb operator. The first part in (3), containing $\langle ij|ab \rangle \langle ab|ij \rangle$, is called direct MP2 energy $E_d^{(2)}$ and the second part, containing $-\langle ij|ab \rangle \langle ab|ji \rangle$, is called exchange MP2 energy $E_x^{(2)}$. Hartree atomic units are used throughout the paper.

A. Reducing the computational cost

In order to give a clear overview of the main strategy to reduce the computational cost of the MP2 energy calculation, we limit ourselves to a schematic description (neglecting spin and periodicity and considering only Γ -point sampling of the Brillouin zone) in this subsection. A proper and comprehensive derivation is given in II B. Furthermore, we restrict to the computationally most expensive part, the exchange term of the MP2 energy (3),

given by

$$E_x^{(2)} = -\frac{1}{2} \sum_{ij}^{\text{occ.}} \sum_{ab}^{\text{virt.}} \frac{\langle ij|ab \rangle \langle ab|ji \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}. \quad (5)$$

In this canonical form the (direct/exchange) MP2 energy has a quintic scaling with the system size. This is due to the fact that the two-electron integrals $\langle ij|ab \rangle \langle ab|ji \rangle$ have to be evaluated for all combinations of i, j, a, b . The computation of $\langle ij|ab \rangle$ for given i, j, a, b scales linear with the system size. This can be seen when the two-electron integrals are written in the plane-wave basis $\{|\mathbf{G}\rangle\}$, $\langle \mathbf{r}|\mathbf{G}\rangle \sim e^{i\mathbf{G}\mathbf{r}}$:

$$\langle ij|ab \rangle = \frac{1}{\Omega} \sum_{\mathbf{G}} \frac{4\pi}{\mathbf{G}^2} \langle i|e^{-\mathbf{G}\hat{\mathbf{r}}}|a\rangle \langle j|e^{\mathbf{G}\hat{\mathbf{r}}}|b\rangle, \quad (6)$$

where Ω is the volume of the system, and the overlap densities are a Fourier transform \mathcal{F} of HF orbitals,

$$\langle i|e^{-\mathbf{G}\hat{\mathbf{r}}}|a\rangle = \int d^3r \varphi_i^*(\mathbf{r}) \varphi_a(\mathbf{r}) e^{-i\mathbf{G}\mathbf{r}} = \mathcal{F}[\varphi_i^* \varphi_a](\mathbf{G}). \quad (7)$$

Our strategy to reduce the computational cost consists of the idea to decouple the sums over the bands i, j, a, b such that the summation over all virtual bands a, b can be performed in advance. The decoupling of the band summation is achieved by a Laplace transformation of the energy denominator [40]:

$$\frac{1}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} = -\int_0^\infty d\tau e^{(\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b)\tau}. \quad (8)$$

It is then possible to rewrite the summand in Eq. (5) such that the sum over a and b can be performed first:

$$-\sum_{ab}^{\text{virt.}} \frac{\langle ij|ab \rangle \langle ab|ji \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad (9)$$

$$\begin{aligned} &= \frac{1}{\Omega^2} \int_0^\infty d\tau e^{(\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b)\tau} \sum_{\mathbf{G}\mathbf{G}'} \frac{4\pi}{\mathbf{G}^2} \frac{4\pi}{\mathbf{G}'^2} \\ &\times \sum_{ab}^{\text{virt.}} \langle i|e^{-\mathbf{G}'\hat{\mathbf{r}}}|a\rangle \langle j|e^{\mathbf{G}'\hat{\mathbf{r}}}|b\rangle \langle a|e^{-\mathbf{G}\hat{\mathbf{r}}}|j\rangle \langle b|e^{\mathbf{G}\hat{\mathbf{r}}}|i\rangle \quad (10) \\ &= \frac{1}{\Omega^2} \int_0^\infty d\tau \sum_{\mathbf{G}\mathbf{G}'} \frac{4\pi}{\mathbf{G}^2} \frac{4\pi}{\mathbf{G}'^2} \\ &\times \langle i|e^{-\mathbf{G}'\hat{\mathbf{r}}} \sum_a^{\text{virt.}} \left[\langle a|e^{-\mathbf{G}\hat{\mathbf{r}}}|j\rangle e^{(\varepsilon_j - \varepsilon_a)\tau} |a\rangle \right] \\ &\times \langle j|e^{+\mathbf{G}'\hat{\mathbf{r}}} \sum_b^{\text{virt.}} \left[\langle b|e^{+\mathbf{G}\hat{\mathbf{r}}}|i\rangle e^{(\varepsilon_i - \varepsilon_b)\tau} |b\rangle \right]. \quad (11) \end{aligned}$$

The squared brackets indicate how the sum over a and b can be performed in advance for each considered \mathbf{G} and τ . This summation defines transformed states:

$$|w_j^{(\mathbf{G}\tau)}\rangle = \sum_a C_{ja}^{(\mathbf{G}\tau)} |a\rangle, \quad (12)$$

where the transformation matrix, for given \mathbf{G} and τ , reads:

$$C_{ja}^{(\mathbf{G}\tau)} = e^{(\varepsilon_j - \varepsilon_a)\tau} \langle a | e^{-\mathbf{G}\hat{\mathbf{r}}} | j \rangle. \quad (13)$$

Then the MP2 exchange energy takes the form of a Fock-like energy:

$$E_x^{(2)} = \frac{1}{2} \frac{1}{\Omega^2} \int_0^\infty d\tau \sum_{\mathbf{G}} \frac{4\pi}{\mathbf{G}^2} \sum_{ij}^{\text{occ.}} \langle ij | w_j^{(\mathbf{G}\tau)} w_i^{(-\mathbf{G},\tau)} \rangle. \quad (14)$$

Since the τ -integration is performed by a quadrature where the number of τ -points is largely independent of the system size, the scaling of the exchange term, $E_x^{(2)}$, is reduced to $\mathcal{O}(N^4)$. For the direct term, $E_d^{(2)}$, the scaling is reduced to $\mathcal{O}(N^3)$ using the same technique. However, the calculation of the transformed states (12) also possesses a quartic scaling such that the algorithm scales with $\mathcal{O}(N^4)$ for both the direct and the exchange MP2 energie.

B. MP2 for periodic systems

In the following subsections we elaborate the aforesaid strategy to reduce the scaling of the MP2 energy for a periodic system in detail. For a periodic system the MP2

energy per unit cell is given [28] by

$$E^{(2)} = \frac{1}{2} \frac{1}{N} \sum_{IJ}^{\text{occ.}} \sum_{AB}^{\text{virt.}} \frac{\langle IJ | AB \rangle [\langle AB | IJ \rangle - \langle AB | JI \rangle]}{\varepsilon_I + \varepsilon_J - \varepsilon_A - \varepsilon_B}. \quad (15)$$

Here N is the number of unit cells composing the entire system. The capital letters I, J, A, B are composite indices representing all quantum numbers. Still I and J run over all occupied Hartree-Fock one-electron states whereas A and B pass through all virtual (unoccupied) states of the system. The ε_I 's and ε_A 's are the occupied and virtual Hartree-Fock one-electron energies respectively.

We assume the system to be periodic (e.g. a periodic 3D lattice). Consequently the quantum numbers are given by a band index, a crystal wave vector and a spin state: $I = (i, \mathbf{k}_1, s_1), J = (j, \mathbf{k}_2, s_2)$, etc. Since the Coulomb operator does not affect the spin degree of freedom the two-electron integrals reduce to

$$\begin{aligned} \langle IJ | AB \rangle &= \langle i\mathbf{k}_1 s_1, j\mathbf{k}_2 s_2 | a\mathbf{k}_3 s_3, b\mathbf{k}_4 s_4 \rangle \\ &= \langle i\mathbf{k}_1 s_1, j\mathbf{k}_2 s_2 | a\mathbf{k}_3 s_1, b\mathbf{k}_4 s_2 \rangle \delta_{s_1 s_3} \delta_{s_2 s_4}. \end{aligned} \quad (16)$$

The spin-unrestricted MP2 energy per unit cell of a periodic system can thus be written as, $E^{(2)} = E_d^{(2)} + E_x^{(2)}$,

$$E_d^{(2)} = \frac{1}{2} \frac{1}{N} \sum_{\mathbf{k}_1 \dots \mathbf{k}_4}^{\text{BZ}} \sum_{ij}^{\text{occ.}} \sum_{ab}^{\text{virt.}} \sum_{ss'}^{\uparrow\downarrow} \frac{|\langle i\mathbf{k}_1 s, j\mathbf{k}_2 s' | a\mathbf{k}_3 s, b\mathbf{k}_4 s' \rangle|^2}{\varepsilon_{i\mathbf{k}_1 s} + \varepsilon_{j\mathbf{k}_2 s'} - \varepsilon_{a\mathbf{k}_3 s} - \varepsilon_{b\mathbf{k}_4 s'}}, \quad (17)$$

$$E_x^{(2)} = -\frac{1}{2} \frac{1}{N} \sum_{\mathbf{k}_1 \dots \mathbf{k}_4}^{\text{BZ}} \sum_{ij}^{\text{occ.}} \sum_{ab}^{\text{virt.}} \sum_s^{\uparrow\downarrow} \frac{\langle i\mathbf{k}_1 s, j\mathbf{k}_2 s | a\mathbf{k}_3 s, b\mathbf{k}_4 s \rangle \langle a\mathbf{k}_3 s, b\mathbf{k}_4 s | j\mathbf{k}_2 s, i\mathbf{k}_1 s \rangle}{\varepsilon_{i\mathbf{k}_1 s} + \varepsilon_{j\mathbf{k}_2 s} - \varepsilon_{a\mathbf{k}_3 s} - \varepsilon_{b\mathbf{k}_4 s}}, \quad (18)$$

where BZ stands for the first Brillouin zone. For brevity, most of the following calculations are performed for the exchange term $E_x^{(2)}$ only. Furthermore the spin-restricted case is assumed, i.e. $\sum_s^{\uparrow\downarrow} \rightarrow 2$, in order to attain a compact notation:

$$\begin{aligned} E_x^{(2)} &= \\ &= -\frac{1}{N} \sum_{\mathbf{k}_1 \dots \mathbf{k}_4}^{\text{BZ}} \sum_{ij}^{\text{occ.}} \sum_{ab}^{\text{virt.}} \frac{\langle i\mathbf{k}_1, j\mathbf{k}_2 | a\mathbf{k}_3, b\mathbf{k}_4 \rangle \langle a\mathbf{k}_3, b\mathbf{k}_4 | j\mathbf{k}_2, i\mathbf{k}_1 \rangle}{\varepsilon_{i\mathbf{k}_1} + \varepsilon_{j\mathbf{k}_2} - \varepsilon_{a\mathbf{k}_3} - \varepsilon_{b\mathbf{k}_4}}. \end{aligned} \quad (19)$$

Note that the two-electron integrals $\langle i\mathbf{k}_1, j\mathbf{k}_2 | a\mathbf{k}_3, b\mathbf{k}_4 \rangle$ are non-vanishing only if $\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{G}$, where \mathbf{G} is some arbitrary reciprocal lattice vector. This can be interpreted as crystal momentum conservation. Hence, the two-electron integrals depend only on three \mathbf{k} -points. In Appendix A the two-electron integrals are written in

the plane-wave basis, revealing this crystal momentum conservation. However, beside this cubic scaling in the number of \mathbf{k} -points the system size scaling is still quintic. In the next step we apply the aforesaid Laplace transform in order to decouple the band summations.

C. Laplace transformed MP2

In (17), (18), and (19) the sums over i, j, a, b can be decoupled by applying a Laplace transformation of the energy denominator [40]:

$$\begin{aligned} &\frac{1}{\varepsilon_{i\mathbf{k}_1} + \varepsilon_{j\mathbf{k}_2} - \varepsilon_{a\mathbf{k}_3} - \varepsilon_{b\mathbf{k}_4}} \\ &= -\int_0^\infty d\tau e^{(\varepsilon_{i\mathbf{k}_1} + \varepsilon_{j\mathbf{k}_2} - \varepsilon_{a\mathbf{k}_3} - \varepsilon_{b\mathbf{k}_4})\tau}. \end{aligned} \quad (20)$$

Note that $\varepsilon_{i\mathbf{k}_1}, \varepsilon_{j\mathbf{k}_2} < \varepsilon_F$ and $\varepsilon_{a\mathbf{k}_3}, \varepsilon_{b\mathbf{k}_4} > \varepsilon_F$ where ε_F is the Fermi energy. Thus the positive definiteness of the denominator and exponent is guaranteed. This leads to the well-known Laplace transformed MP2 (LTMP2) expression [40]:

$$E_x^{(2)} = \frac{1}{N} \int_0^\infty d\tau \sum_{\mathbf{k}_1 \dots \mathbf{k}_4}^{\text{BZ}} \sum_{ij}^{\text{occ.}} \sum_{ab}^{\text{virt.}} \langle i\mathbf{k}_1, j\mathbf{k}_2 | a\mathbf{k}_3, b\mathbf{k}_4 \rangle \times \langle a\mathbf{k}_3, b\mathbf{k}_4 | j\mathbf{k}_2, i\mathbf{k}_1 \rangle e^{(\varepsilon_{i\mathbf{k}_1} + \varepsilon_{j\mathbf{k}_2} - \varepsilon_{a\mathbf{k}_3} - \varepsilon_{b\mathbf{k}_4})\tau}. \quad (21)$$

Although the decoupling comes at the price of an additional integration, it has no effect to the scaling. Due to the exponentially decreasing behavior in τ , the integration can be performed by a quadrature [41, 42] using only a few τ -points.

As already indicated in Sec. II A, the summations over the virtual bands a, b can now be performed in advance. In the following we will derive an analogous decomposition to Eq. (12) for periodic system.

D. LTMP2 in the plane-wave basis and presummation over all virtual bands

The algorithm presented here is implemented in VASP which uses a plane-wave basis set. Hence the two-electron integrals $\langle i\mathbf{k}_1, j\mathbf{k}_2 | a\mathbf{k}_3, b\mathbf{k}_4 \rangle$ will be evaluated in the plane-wave basis. The orbitals obey Bloch's theorem due to the periodicity of the system. In position representation they can be written as

$$\varphi_{i\mathbf{k}}(\mathbf{r}) = \langle \mathbf{r} | i\mathbf{k} \rangle = \frac{1}{\sqrt{\Omega}} e^{i\mathbf{k}\mathbf{r}} u_{i\mathbf{k}}(\mathbf{r}), \quad (22)$$

where $u_{i\mathbf{k}}$ is the cell periodic part of the orbital. The system is decomposed into N unit cells of volume Ω_0 such that the volume of the entire system is $\Omega = N\Omega_0$. At the boundaries the Born-von Karman conditions are assumed. The states are normalized by

$$\langle i\mathbf{k}_1 | j\mathbf{k}_2 \rangle = \int_\Omega d^3r \varphi_{i\mathbf{k}_1}^*(\mathbf{r}) \varphi_{j\mathbf{k}_2}(\mathbf{r}) = \delta_{ij} \delta_{\mathbf{k}_1 \mathbf{k}_2}, \quad (23)$$

which implies

$$\int_{\Omega_0} d^3r u_{i\mathbf{k}}^*(\mathbf{r}) u_{j\mathbf{k}}(\mathbf{r}) = \Omega_0 \delta_{ij}. \quad (24)$$

Using Eq. (22) and expanding the Coulomb kernel, $1/|\mathbf{r} - \mathbf{r}'|$, in Fourier space, the two-electron integrals can be written as:

$$\begin{aligned} & \langle i\mathbf{k}_1, j\mathbf{k}_2 | a\mathbf{k}_3, b\mathbf{k}_4 \rangle \\ &= \frac{1}{\Omega} \delta_{T(\mathbf{k}_1 - \mathbf{k}_3), T(\mathbf{k}_4 - \mathbf{k}_2)} \sum_{\mathbf{G}} \frac{4\pi}{[\mathbf{G} + T(\mathbf{k}_1 - \mathbf{k}_3)]^2} \\ & \times \langle i\mathbf{k}_1 | e^{+i[\mathbf{G} + T(\mathbf{k}_1 - \mathbf{k}_3)]\hat{\mathbf{r}}} | a\mathbf{k}_3 \rangle_{\Omega_0} \\ & \times \langle j\mathbf{k}_2 | e^{-i[\mathbf{G} + T(\mathbf{k}_4 - \mathbf{k}_2)]\hat{\mathbf{r}}} | b\mathbf{k}_4 \rangle_{\Omega_0}. \end{aligned} \quad (25)$$

A step-by-step derivation can be found in Appendix A. An explanation of the notation is in order: $\sum_{\mathbf{G}}$ is a sum over all (infinitely many) reciprocal lattice vectors \mathbf{G} arising from the real lattice of the system. $T(\mathbf{k})$ is a function that maps \mathbf{k} back to the first BZ along a translation by an appropriate reciprocal lattice vector. In analogy to (7) the overlap densities are defined by

$$\langle i\mathbf{k}_1 | e^{-i\mathbf{G}\hat{\mathbf{r}}} | a\mathbf{k}_3 \rangle_{\Omega_0} := N \int_{\Omega_0} d^3r \varphi_{i\mathbf{k}_1}^*(\mathbf{r}) \varphi_{a\mathbf{k}_3}(\mathbf{r}) e^{-i\mathbf{G}\mathbf{r}}. \quad (26)$$

Note that this unitless quantity does not explicitly depend on N , since N is balanced by the normalization factors of the orbitals (22). However, there is an implicit dependence since different N lead to different Born-von Karman boundaries, hence a different mesh of crystal wave vectors (\mathbf{k} -point mesh). Inserting (25) into (21), one sum over the BZ, here $\mathbf{k}_4 \rightarrow T(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3)$, can be eliminated. Note that the Kronecker deltas for the \mathbf{k} -vectors which occur in $\langle i\mathbf{k}_1, j\mathbf{k}_2 | a\mathbf{k}_3, b\mathbf{k}_4 \rangle$ and $\langle a\mathbf{k}_3, b\mathbf{k}_4 | j\mathbf{k}_2, i\mathbf{k}_1 \rangle$ are equivalent. A substitution $\mathbf{q} = T(\mathbf{k}_2 - \mathbf{k}_3)$ then leads to

$$\begin{aligned} E_x^{(2)} &= \frac{1}{\Omega_0^2} \frac{1}{N^3} \int_0^\infty d\tau \sum_{\mathbf{k}_1 \mathbf{k}_2 \mathbf{q}}^{\text{BZ}} \sum_{ij}^{\text{occ.}} \sum_{\mathbf{G}}^{\text{virt.}} \frac{4\pi}{[\mathbf{G} + T(\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{q})]^2} \sum_{\mathbf{G}'} \frac{4\pi}{(\mathbf{G}' - \mathbf{q})^2} \\ & \times \sum_a^{\text{virt.}} \langle i\mathbf{k}_1 | e^{+i[\mathbf{G} + T(\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{q})]\hat{\mathbf{r}}} | aT(\mathbf{k}_2 - \mathbf{q}) \rangle_{\Omega_0} \times \langle aT(\mathbf{k}_2 - \mathbf{q}) | e^{+i(\mathbf{G}' - \mathbf{q})\hat{\mathbf{r}}} | j\mathbf{k}_2 \rangle_{\Omega_0} e^{(\varepsilon_{j\mathbf{k}_2} - \varepsilon_{aT(\mathbf{k}_2 - \mathbf{q})})\tau} \\ & \times \sum_b^{\text{virt.}} \langle j\mathbf{k}_2 | e^{-i[\mathbf{G} + T(\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{q})]\hat{\mathbf{r}}} | bT(\mathbf{k}_1 + \mathbf{q}) \rangle_{\Omega_0} \times \langle bT(\mathbf{k}_1 + \mathbf{q}) | e^{-i(\mathbf{G}' - \mathbf{q})\hat{\mathbf{r}}} | i\mathbf{k}_1 \rangle_{\Omega_0} e^{(\varepsilon_{i\mathbf{k}_1} - \varepsilon_{bT(\mathbf{k}_1 + \mathbf{q})})\tau}. \end{aligned} \quad (27)$$

The MP2 energy per unit cell does not explicitly depend on N , although $1/N^3$ appears in the above formula.

Again the N dependence is only implicit since N defines

the density of the allowed \mathbf{k} -points. This becomes evident when we perform the thermodynamic limit, $N \rightarrow \infty$, $\Omega/N = \Omega_0 = \text{const.}$. The density of the crystal wave vectors \mathbf{k} becomes infinite and sums turn into integrals:

$$\frac{1}{N} \sum_{\mathbf{k}} \dots \longrightarrow \Omega_0 \int_{\text{BZ}} \frac{d^3 k}{(2\pi)^3} \dots, \quad (28)$$

which implies:

$$\frac{1}{\Omega_0^2} \frac{1}{N^3} \sum_{\mathbf{k}_1 \mathbf{k}_2 \mathbf{q}} \dots \longrightarrow \Omega_0 \int_{(\text{BZ})^3} \frac{d^3 k_1}{(2\pi)^3} \frac{d^3 k_2}{(2\pi)^3} \frac{d^3 q}{(2\pi)^3} \dots \quad (29)$$

Since $\Omega \rightarrow \infty$ the normalization factor has to be dropped in (22) and we write $\varphi_{i\mathbf{k}}(\mathbf{r}) \longrightarrow e^{i\mathbf{k}\mathbf{r}} u_{i\mathbf{k}}(\mathbf{r})$ such that the overlap densities (26) now read:

$$\langle i\mathbf{k}_1 | e^{-i\mathbf{G}'\mathbf{r}} | a\mathbf{k}_3 \rangle_{\Omega_0} \longrightarrow \frac{1}{\Omega_0} \int_{\Omega_0} d^3 r \varphi_{i\mathbf{k}_1}^*(\mathbf{r}) \varphi_{a\mathbf{k}_3}(\mathbf{r}) e^{-i\mathbf{G}'\mathbf{r}}. \quad (30)$$

The formula (27) already suggests a possibility to perform a presummation over all virtual bands a, b : For all \mathbf{G}', \mathbf{q} and τ define a decomposition by

$$|w_{i\mathbf{k}}^{(\mathbf{G}'\mathbf{q}\tau)}\rangle = \sum_b^{\text{virt.}} C_{ib,\mathbf{k}}^{(\mathbf{G}'\mathbf{q}\tau)} |b\mathbf{k}\rangle. \quad (31)$$

The coefficients are defined as:

$$C_{ib,\mathbf{k}}^{(\mathbf{G}'\mathbf{q}\tau)} = e^{(\varepsilon_{iT(\mathbf{k}-\mathbf{q})} - \varepsilon_{b\mathbf{k}})\tau} \times \langle b\mathbf{k} | e^{-i(\mathbf{G}'-\mathbf{q})\mathbf{r}} | iT(\mathbf{k}-\mathbf{q}) \rangle_{\Omega_0}. \quad (32)$$

The decomposition (31) can be performed in advance such that for a given \mathbf{G}', \mathbf{q} and τ the large number of virtual bands are reduced to transformed states $|w_{i\mathbf{k}}^{(\mathbf{G}'\mathbf{q}\tau)}\rangle$ labeled by the few occupied indices i . Hence the MP2 energy per unit cell can be written in a form which involves summations only over the occupied indices i, j :

$$\begin{aligned} E_x^{(2)} = & \Omega_0 \int_0^\infty d\tau \int_{(\text{BZ})^3} \frac{d^3 q}{(2\pi)^3} \frac{d^3 k_1}{(2\pi)^3} \frac{d^3 k_2}{(2\pi)^3} \\ & \times \sum_{\mathbf{G}\mathbf{G}'} \frac{4\pi}{[\mathbf{G} + T(\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{q})]^2} \frac{4\pi}{(\mathbf{G}' - \mathbf{q})^2} \\ & \times \sum_{ij}^{\text{occ.}} \left\langle i\mathbf{k}_1 \left| e^{+i[\mathbf{G}+T(\mathbf{k}_1-\mathbf{k}_2+\mathbf{q})]\mathbf{r}} \right| w_{jT(\mathbf{k}_2-\mathbf{q})}^{(-\mathbf{G}',-\mathbf{q},\tau)} \right\rangle_{\Omega_0} \\ & \times \left\langle j\mathbf{k}_2 \left| e^{-i[\mathbf{G}+T(\mathbf{k}_1-\mathbf{k}_2+\mathbf{q})]\mathbf{r}} \right| w_{iT(\mathbf{k}_1+\mathbf{q})}^{(+\mathbf{G}',+\mathbf{q},\tau)} \right\rangle_{\Omega_0}. \end{aligned} \quad (33)$$

E. Time reversal symmetry

For a more convenient implementation in a computer code it is advantageous to exploit the time reversal symmetry. With its aid we can turn both overlap densities

in Eq. (33) into the same form, i.e. to avoid mixtures of $+\mathbf{G}'$ and $-\mathbf{G}'$ as well as mixtures of $+\mathbf{G}'$ and $-\mathbf{G}'$. This can be done in the following way. If no external magnetic field is applied and spin-orbit coupling is ignored the Hartree-Fock orbitals (22) obey time reversal symmetry: $\varphi_{i\mathbf{k}}^* = \varphi_{i-\mathbf{k}}$ and $\varepsilon_{i\mathbf{k}} = \varepsilon_{i-\mathbf{k}}$. If we apply this time reversal symmetry to the coefficients (32) we obtain:

$$\left(C_{ib,+\mathbf{k}}^{(+\mathbf{G}',+\mathbf{q},\tau)} \right)^* = C_{ib,-\mathbf{k}}^{(-\mathbf{G}',-\mathbf{q},\tau)}. \quad (34)$$

Hence for (31) we find:

$$\left| \left(w_{i,+\mathbf{k}}^{(+\mathbf{G}',+\mathbf{q},\tau)} \right)^* \right\rangle = \left| w_{i,-\mathbf{k}}^{(-\mathbf{G}',-\mathbf{q},\tau)} \right\rangle. \quad (35)$$

This relation can be applied to the overlap densities in (33). Consider, e.g.,

$$\begin{aligned} & \left\langle i\mathbf{k}_1 \left| e^{+i[\mathbf{G}+T(\mathbf{k}_1-\mathbf{k}_2+\mathbf{q})]\mathbf{r}} \right| w_{jT(\mathbf{k}_2-\mathbf{q})}^{(-\mathbf{G}',-\mathbf{q},\tau)} \right\rangle_{\Omega_0} \\ & = \left\langle w_{jT(\mathbf{k}_2-\mathbf{q})}^{(-\mathbf{G}',-\mathbf{q},\tau)} \left| e^{-i[\mathbf{G}+T(\mathbf{k}_1-\mathbf{k}_2+\mathbf{q})]\mathbf{r}} \right| i\mathbf{k}_1 \right\rangle_{\Omega_0}^* \\ & = \left\langle (i\mathbf{k}_1)^* \left| e^{-i[\mathbf{G}+T(\mathbf{k}_1-\mathbf{k}_2+\mathbf{q})]\mathbf{r}} \right| w_{jT(-\mathbf{k}_2+\mathbf{q})}^{(\mathbf{G}'\mathbf{q}\tau)} \right\rangle_{\Omega_0}^* \end{aligned} \quad (36)$$

In a last step we substitute $\mathbf{k}_2 \rightarrow -\mathbf{k}_2$ in (33) such that we find a convenient and more "symmetric" formula for the MP2 exchange energy per unit cell:

$$\begin{aligned} E_x^{(2)} = & \Omega_0 \int_0^\infty d\tau \int_{(\text{BZ})^3} \frac{d^3 q}{(2\pi)^3} \frac{d^3 k_1}{(2\pi)^3} \frac{d^3 k_2}{(2\pi)^3} \\ & \times \sum_{\mathbf{G}\mathbf{G}'} \frac{4\pi}{[\mathbf{G} + T(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{q})]^2} \frac{4\pi}{(\mathbf{G}' - \mathbf{q})^2} \\ & \times \sum_{ij}^{\text{occ.}} \left\langle (i\mathbf{k}_1)^* \left| e^{-i[\mathbf{G}+T(\mathbf{k}_1+\mathbf{k}_2+\mathbf{q})]\mathbf{r}} \right| w_{jT(\mathbf{k}_2+\mathbf{q})}^{(\mathbf{G}'\mathbf{q},\tau)} \right\rangle_{\Omega_0}^* \\ & \times \left\langle (j\mathbf{k}_2)^* \left| e^{-i[\mathbf{G}+T(\mathbf{k}_1+\mathbf{k}_2+\mathbf{q})]\mathbf{r}} \right| w_{iT(\mathbf{k}_1+\mathbf{q})}^{(\mathbf{G}'\mathbf{q}\tau)} \right\rangle_{\Omega_0}. \end{aligned} \quad (37)$$

As for the exchange term the same procedure can be applied to the direct term. Starting with (17) one finds:

$$\begin{aligned} E_d^{(2)} = & \Omega_0 \int_0^\infty d\tau \int_{\text{BZ}} \frac{d^3 q}{(2\pi)^3} \sum_{\mathbf{G}\mathbf{G}'} \frac{4\pi}{(\mathbf{G} + \mathbf{q})^2} \frac{4\pi}{(\mathbf{G}' - \mathbf{q})^2} \\ & \times \left| \int_{\text{BZ}} \frac{d^3 k}{(2\pi)^3} \sum_i^{\text{occ.}} \left\langle i\mathbf{k} \left| e^{-i(\mathbf{G}+\mathbf{q})\mathbf{r}} \right| w_{iT(\mathbf{k}+\mathbf{q})}^{(\mathbf{G}'\mathbf{q}\tau)} \right\rangle_{\Omega_0} \right|^2. \end{aligned} \quad (38)$$

III. IMPLEMENTATION

The presented LTMP2 method is implemented in the Vienna ab-initio simulation package (VASP) [31, 32] based on the projector augmented wave (PAW) [43] method. For brevity this section restricts to the Γ -only version, i.e. we neglect \mathbf{k} -point sampling of the Brillouin zone here.

A. General strategy

The four major steps of the algorithm can be summarized as follows:

- (i) Compute and store all overlap densities (7).
- (ii) Loop over all τ -points and reciprocal vectors \mathbf{G} of the outer loops in Eq. (14).
- (iii) Calculate the transformation matrix (13) for this τ and \mathbf{G} using the stored overlap densities of step (i) and construct the transformed states $|w_i\rangle$, see (12).
- (iv) Perform
 - (a) a Hartree-like routine,
 - (b) a Fock-like routine

to calculate the direct and exchange MP2 contribution for this τ and \mathbf{G} , i.e. evaluate the two-electron integral $\langle ij|w_j w_i\rangle$ and sum over i, j in (14).

The outer loop over plane-waves \mathbf{G} is limited by an adjustable plane-wave cutoff. The outer loop over the τ -points is performed by a quadrature [42]. Figure 1 shows the pseudocode for the serial Γ -only implementation of the algorithm.

B. Parallelization

The outer \mathbf{G} -loop (see step (ii) in Sec. III A or Eq. 14) provides a powerful approach to parallelize the algorithm. The entire set of reciprocal lattice vectors \mathcal{G} can be divided into $\mathcal{N}_{\mathbf{G}}$ independent subsets. This leads to a very high parallelization efficiency as long as the total number of reciprocal lattice vectors $N_{\mathbf{G}}$ is larger than $\mathcal{N}_{\mathbf{G}}$. Furthermore, on a second level, a parallelization is implemented for the evaluation of the sum over occupied bands: The set of occupied bands \mathcal{I} is divided into $\mathcal{N}_{\mathbf{B}}$ subsets such that one band summation (say over j) can be calculated in parallel. Pseudocode for this strategy can be found in Fig. 8.

C. Formal system size scaling: computation time and memory

Calculating the exchange term (which has the steepest scaling) results in a computation time that scales with the fourth power of the system size. This stems from the fact that for every combination of \mathbf{G}, i, j fast Fourier transforms (FFT) have to be performed. In a spin-unrestricted calculation, the computation time will be twice as large as for a spin-restricted case. The number of τ -points for the quadrature, to numerically evaluate the τ -integration, is largely independent of the system size. Table I shows the formal scaling for each step.

```

# overlap densities (i)
for all  $i, a$  do
   $\rho_{ai}(\mathbf{r}) \leftarrow \varphi_a^*(\mathbf{r})\varphi_i(\mathbf{r})$ 
   $\tilde{\rho}_{ai} \leftarrow \text{FFT}[\rho_{ai}]$ 
  store  $\tilde{\rho}_{ai}$ 
end for

for all  $\tau$  do
  for all  $\mathbf{G}$  do

    # transformed states (iii)
    for all  $i$  do
      for all  $a$  do
         $C_{ia} \leftarrow e^{(\varepsilon_i - \varepsilon_a)\tau} \tilde{\rho}_{ai}(\mathbf{G})$ 
      end for
       $\tilde{w}_i \leftarrow \sum_a C_{ia} \tilde{\varphi}_a$ 
      store  $\tilde{w}_i$ 
    end for

    # direct term (iv.a)
     $w_i \leftarrow \text{FFT}^{-1}[\tilde{w}_i]$  for all  $i$ 
     $\varrho(\mathbf{r}) \leftarrow \sum_i \varphi_i^*(\mathbf{r})w_i(\mathbf{r})$ 
     $\tilde{\varrho} \leftarrow \text{FFT}[\varrho]$ 
     $e_d^{(2)} \leftarrow e_d^{(2)} + \sum_{\mathbf{G}'} \frac{4\pi}{G'^2} |\tilde{\varrho}(\mathbf{G}')|^2$ 

    # exchange term (iv.b)
     $w_i \leftarrow \text{FFT}^{-1}[\tilde{w}_i]$  for all  $i$ 
    for all  $i, j$  ( $i \leq j$ ) do
       $\varrho_1(\mathbf{r}) \leftarrow \varphi_i^*(\mathbf{r})w_j(\mathbf{r})$ 
       $\varrho_2(\mathbf{r}) \leftarrow \varphi_j^*(\mathbf{r})w_i(\mathbf{r})$ 
       $\tilde{\varrho}_1 \leftarrow \text{FFT}[\varrho_1]$ 
       $\tilde{\varrho}_2 \leftarrow \text{FFT}[\varrho_2]$ 
       $e_x^{(2)} \leftarrow e_x^{(2)} + \sum_{\mathbf{G}'} \frac{4\pi}{G'^2} \tilde{\varrho}_1^*(\mathbf{G}')\tilde{\varrho}_2(\mathbf{G}')$ 
    end for
     $E_d^{(2)} \leftarrow E_d^{(2)} + \frac{4\pi}{G^2} e_d^{(2)}$ 
     $E_x^{(2)} \leftarrow E_x^{(2)} + \frac{4\pi}{G^2} e_x^{(2)}$ 
  end for
end for

```

Figure 1. Pseudocode of the serial Γ -only implementation. Hartree-Fock orbitals φ_i, φ_a and energies $\varepsilon_i, \varepsilon_a$ for all occupied and virtual states are assumed.

Regarding the memory consumption only the orbitals and the overlap densities (7) are of relevance. The latter has the steepest scaling: $N_i N_a N_{\text{FFT}}$. The memory requirement of the orbitals scales only quadratically since each orbital has to be stored on the entire FFT grid: $(N_i + N_a) N_{\text{FFT}}$. Regarding \mathbf{k} -point sampling of the Brillouin zone, a linear dependence of $N_{\mathbf{k}}$ has to be multiplied in both cases. For a spin-unrestricted calculation the memory requirement doubles.

D. Internal cutoff extrapolation

Similar to other MP2 implementations or RPA codes the LTMP2 algorithm converges slowly with respect to the number of basis functions (plane-wave cutoff). However, within the presented algorithm an internal extrap-

step	system size	k-points
(i)	$N_i N_a N_{\text{FFT}}$	N_k^2
(iii)	$N_i N_a N_G^2$	N_k^2
(iv.a)	$N_i N_G N_{\text{FFT}}$	N_k^2
(iv.b)	$N_i^2 N_G N_{\text{FFT}}$	N_k^3

Table I. Formal scaling of the computation time for the different steps of the algorithm. The steps are described in Sec. III A. N_i and N_a are the number of occupied and virtual bands respectively. N_G is the number of reciprocal lattice vectors, N_{FFT} the number of FFT grid points, and N_k is the number of \mathbf{k} -points.

olation of the plane-wave cutoff could be implemented comfortably with negligible influence on the computation time. Again, as for the parallelization, the outer \mathbf{G} -loop can be exploited: MP2 energies are computed for different cutoff spheres by calculating intermediate results of the outer \mathbf{G} -loop on the fly. In this way a certain number of MP2 energies are calculated for cutoffs reaching from 70% to 100% of the user-given plane-wave cutoff E_{cut} . These MP2 energies are then extrapolated using the known asymptotic behavior for large E_{cut} [44, 45]:

$$E^{(2)}(E_{\text{cut}}) - E^{(2)}(\infty) \sim E_{\text{cut}}^{-3/2}. \quad (39)$$

The reliability of this extrapolation depends on the user-given E_{cut} since the asymptotic behavior is strictly true only for $E_{\text{cut}} \rightarrow \infty$.

IV. BENCHMARK CALCULATIONS

In order to show the potential of the new LTMP2 method we performed several benchmark calculations. Computations on supercells of solid lithium hydride (LiH) served as a benchmark to demonstrate the parallelization efficiency and system size scaling. The advantage of an internal plane-wave cutoff extrapolation is shown by means of binding energy calculations of methane (CH_4) in a chabazite crystal ($\text{AlHO}_{24}\text{Si}_{11}$). The calculations were performed with VASP in which the new LTMP2 code was implemented. For each benchmark calculation shown in this chapter we used a Γ -only and spin-restricted setting. Regarding the τ -integration 6 τ -points turned out to be accurate enough (< 1 meV convergence) for all considerations in this section.

A. Measured parallelization efficiency

To demonstrate the parallelization efficiency we computed the MP2 energy of solid LiH using a supercell containing $4 \times 4 \times 4$ primitive cells and 128 atoms. The computation time was measured against the number of cores. A plane-wave cutoff of 289 eV lead to 5600 reciprocal lattice vectors for the outer \mathbf{G} -loop. The total number of orbitals was set to 20480 whereas the number

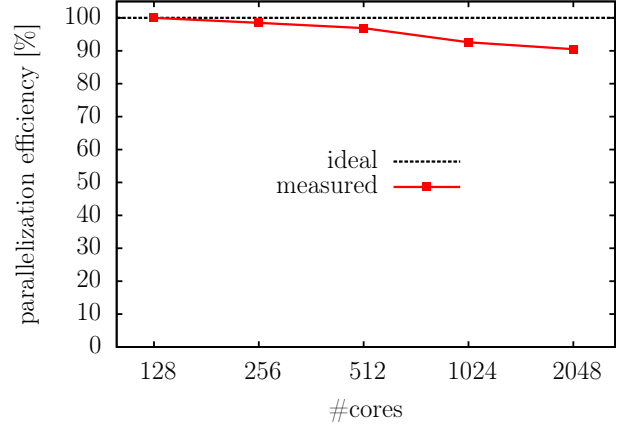


Figure 2. Parallelization efficiency of the LTMP2 calculation of a $4 \times 4 \times 4$ supercell of solid LiH. The efficiency was calculated with reference to the computation time t_{128} of a run using 128 cores: $t_{128}/t_{\text{\#cores}}/(\text{\#cores}/128)$. Note that the abscissa has a logarithmic scale.

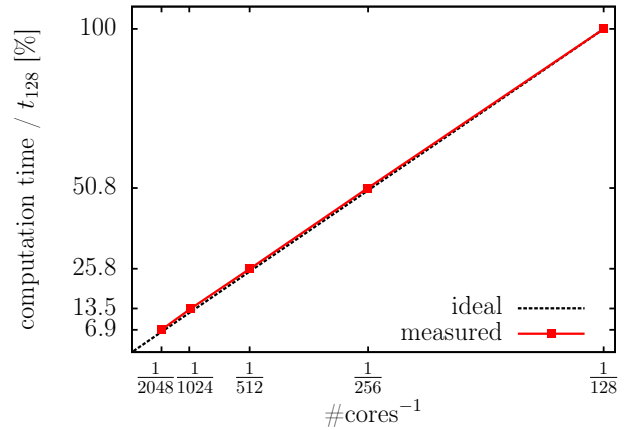


Figure 3. Strong scaling of the LTMP2 calculation of a $4 \times 4 \times 4$ supercell of solid LiH. Here t_{128} is the computation time of a run with 128 cores. In the case of 2048 cores the computation time is reduced to 6.9% of t_{128} which is close to the ideal case of $128/2048 = 6.25\%$.

of occupied orbitals amounted to 128. In each run the number of plane-wave groups was set to $\mathcal{N}_G = \text{\#cores}/4$ which implies $\mathcal{N}_B = 4$ for the band parallelization. Figures 2 and 3 show the parallelization efficiency and the strong scaling. The strong scaling shows that with increasing number of cores the computation time approaches zero as long as the number of reciprocal lattice vectors, N_G , is divisible by the number of plane-wave groups, \mathcal{N}_G .

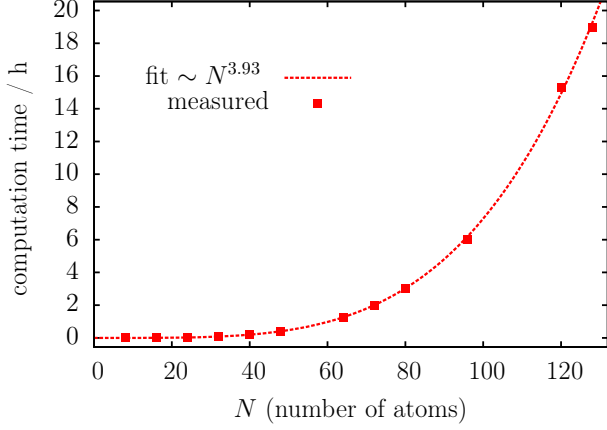


Figure 4. System size scaling of the LTMP2 computation time for various supercells of solid LiH on a modest system with 64 cores.

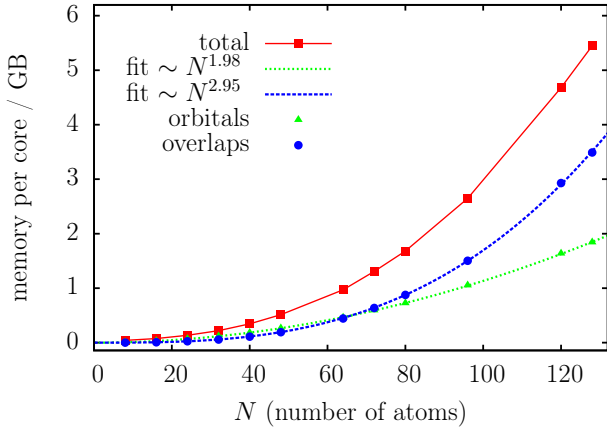


Figure 5. Scaling of the used memory per core with respect to the system size. The two most relevant quantities are the orbitals and the overlap densities (7).

B. Measured system size scaling: computation time and memory

To evaluate the time and memory scaling of the LTMP2 algorithm with respect to the system size, the computation time and the memory consumption was measured against the number of atoms for various supercells of solid LiH. The smallest system containing 8 atoms corresponds to the conventional unit cell of LiH. This cell was replicated in numerous ways to form supercells containing up to 128 atoms. The plane-wave cutoff was set to 289 eV. The calculations were performed on a modest number of 64 CPUs. For the parallelization we used $\mathcal{N}_G = 32$ and $\mathcal{N}_B = 2$. Figure 4 and 5 show the measured scaling of the computation time and memory. The measured scaling exponents match the predicted values.

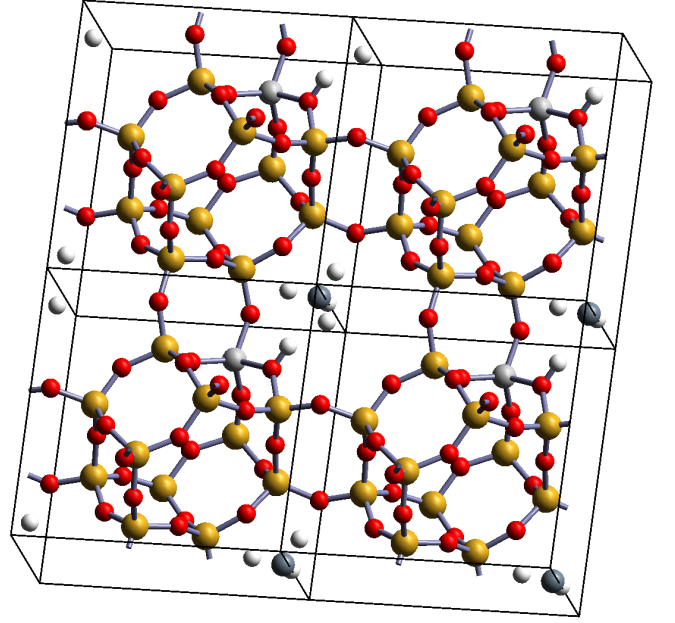


Figure 6. Four unit cells of the chabazite crystal with adsorbed methane molecules. The color code reads: Al (light gray), C (dark gray), H (white), O (red), Si (yellow).

C. Internal cutoff extrapolation

The internal cutoff extrapolation described in Sec. IIID is illustrated by an adsorption energy calculation of CH_4 in a chabazite crystal: the adsorption energy was computed against the plane-wave cutoff E_{cut} . Furthermore, to show that the new LTMP2 implementation gives correct MP2 energies, the existing quintic scaling MP2 code [33] in vasp is taken as a reference. The lattice parameters and the orientation of the CH_4 molecule in a chabazite cage were taken from [21] and then reoptimized using the optB88-vdW functional [46]. The entire system consists of 42 atoms in a unit cell of about 810 \AA^3 . A visualization can be found in Fig. 6. The total number of bands was set to 15872 where the first 100 are occupied bands. For the orbitals a large plane-wave cutoff (ENCUT flag in VASP) of 800 eV was used, and all virtual orbitals were included in the subsequent calculations (not to be confused with the plane-wave cutoff, E_{cut} (ENCUTGW flag in VASP), for the two-electron integrals in the MP2 calculation). The previous MP2 code clearly shows the mentioned $E_{\text{cut}}^{-3/2}$ behavior (39) as can be seen in Fig. 7. It is worth mentioning that the MP2 curve reaches the converged value not before a cutoff of 1000 eV, if a tolerance of 1% is assumed. However, the new LTMP2 algorithm attains this accuracy already at a cutoff of about 300 eV due to the internal extrapolation.

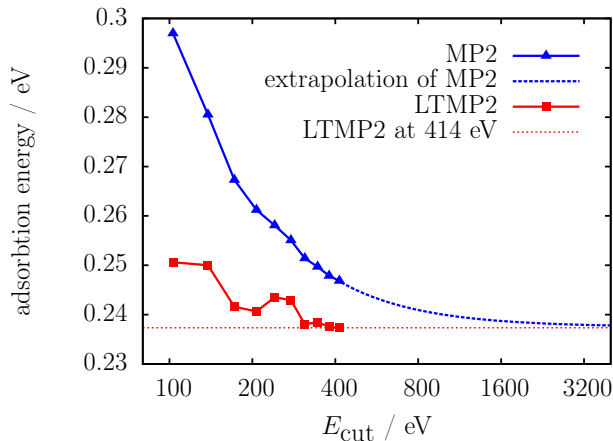


Figure 7. Adsorption energy of methane in a chabazite cage as a function of the plane-wave cutoff E_{cut} . MP2 refers to the previous implementation whereas LTMP2 is the algorithm of this work. For the extrapolation (dashed line) only the last 6 points were taken into account using (39). Note the logarithmic scaling of the abscissa.

V. CONCLUSION AND OUTLOOK

We have presented a new algorithm to calculate the exact MP2 energy for periodic systems, scaling only with the fourth power of the system size, $\mathcal{O}(N^4)$. The lower scaling is a consequence of a Laplace transformed energy denominator of the traditional MP2 formulation. In doing so the summations over the virtual bands can be carried out first, leading to transformed states in dependence of a plane-wave index and a τ -point. The loop over these plane-waves is an outer loop that can be distributed over the CPUs without communication, leading to a very high parallelization efficiency. We showed that the parallelization is extremely close to the ideal case as long as the number of plane-waves can be divided by the number, \mathcal{N}_G , of parallelized plane-wave groups.

The slow convergence of the MP2 energy with respect to the number of basis functions is treated by an extrapolation to an infinite basis set using the exact asymptotic cutoff behavior. In a comparison with the previous MP2 code in VASP, we demonstrated that this extrapolation leads to converged MP2 energies using significantly smaller basis sets (smaller cutoffs).

In future the presented approach could be adapted to more involved electronic correlation energy methods, like second-order screened exchange (SOSEX) [47, 48] or particle-hole ladder diagrams, in order to obtain a similar low complexity. Hence, the presented method can be considered as a step towards systematically improved correlation energies.

Also, calculating interatomic forces should be possible using the presented low-complexity concept, as the self-energy at MP2 level can be obtained along the same lines presented here for the MP2 energy. With the respective

self-energy at hand, the recently published Green's function approach allows to efficiently calculate interatomic forces for perturbative methods, in this case MP2 [49].

VI. ACKNOWLEDGEMENTS

Computations were performed on the Vienna Scientific Cluster, VSC3. We also thank Jiří Klimeš for providing the optimized chabazite structure.

VII. APPENDIX

Appendix A: Two-electron integrals in the plane-wave basis

For writing out a two-electron integral,

$$\begin{aligned} \langle i\mathbf{k}_1, j\mathbf{k}_2 | a\mathbf{k}_3, b\mathbf{k}_4 \rangle \\ = \int d^3r \int d^3r' \frac{\varphi_{i\mathbf{k}_1}^*(\mathbf{r}) \varphi_{j\mathbf{k}_2}^*(\mathbf{r}') \varphi_{a\mathbf{k}_3}(\mathbf{r}) \varphi_{b\mathbf{k}_4}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \end{aligned} \quad (\text{A1})$$

in the plane-wave basis we expand the Coulomb kernel, $1/|\mathbf{r} - \mathbf{r}'|$, in Fourier space:

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \frac{1}{\Omega} \sum_{\mathbf{G}} \sum_{\mathbf{k}}^{\text{BZ}} \frac{4\pi}{(\mathbf{G} + \mathbf{k})^2} e^{i(\mathbf{G} + \mathbf{k})(\mathbf{r} - \mathbf{r}')} . \quad (\text{A2})$$

Note that in the strict sense this is only an approximation. However, in the thermodynamic limit, $N \rightarrow \infty$, $\Omega/N = \Omega_0 = \text{const.}$, the sum $1/\Omega \sum_{\mathbf{k}}$ turns into an integral $\int d^3k/(2\pi)^3$ and (A2) becomes exact. Furthermore we use the identity

$$\sum_{\mathbf{R}} e^{\pm i(\mathbf{k} - \mathbf{k}')\mathbf{R}} = N \delta_{T(\mathbf{k}), T(\mathbf{k}')} . \quad (\text{A3})$$

Here $\sum_{\mathbf{R}}$ is a sum over all N lattice vectors \mathbf{R} of the system and \mathbf{k}, \mathbf{k}' are crystal wave-vectors (not necessarily restricted to the first BZ). The function $T(\mathbf{k})$ is defined on page 4. Inserting (A2) into (A1), we split the integration over the entire system Ω into integrations over unit

cells Ω_0 translated by all lattice vectors:

$$\begin{aligned}
& \langle i\mathbf{k}_1, j\mathbf{k}_2 | a\mathbf{k}_3, b\mathbf{k}_4 \rangle \\
&= \\
& \frac{1}{\Omega} \sum_{\mathbf{R}\mathbf{R}'} \int_{\Omega_0} d\mathbf{r} \int_{\Omega_0} d\mathbf{r}' \sum_{\mathbf{G}} \sum_{\mathbf{k}}^{\text{BZ}} \frac{4\pi}{(\mathbf{G} + \mathbf{k})^2} e^{i(\mathbf{G} + \mathbf{k})(\mathbf{r} + \mathbf{R} - \mathbf{r}' - \mathbf{R}')} \\
& \times e^{-i(\mathbf{k}_1 - \mathbf{k}_3)\mathbf{R}} e^{-i(\mathbf{k}_2 - \mathbf{k}_4)\mathbf{R}'} \varphi_{i\mathbf{k}_1}^*(\mathbf{r}) \varphi_{j\mathbf{k}_2}^*(\mathbf{r}') \varphi_{a\mathbf{k}_3}(\mathbf{r}) \varphi_{b\mathbf{k}_4}(\mathbf{r}') \\
&= \\
& \frac{N^2}{\Omega} \int_{\Omega_0} d\mathbf{r} \int_{\Omega_0} d\mathbf{r}' \sum_{\mathbf{G}} \sum_{\mathbf{k}}^{\text{BZ}} \frac{4\pi}{(\mathbf{G} + \mathbf{k})^2} e^{i(\mathbf{G} + \mathbf{k})(\mathbf{r} - \mathbf{r}')} \\
& \times \delta_{\mathbf{k}, T(\mathbf{k}_1 - \mathbf{k}_3)} \delta_{\mathbf{k}, T(\mathbf{k}_4 - \mathbf{k}_2)} \varphi_{i\mathbf{k}_1}^*(\mathbf{r}) \varphi_{j\mathbf{k}_2}^*(\mathbf{r}') \varphi_{a\mathbf{k}_3}(\mathbf{r}) \varphi_{b\mathbf{k}_4}(\mathbf{r}') \\
&= \\
& \frac{1}{\Omega} \delta_{T(\mathbf{k}_1 - \mathbf{k}_3), T(\mathbf{k}_4 - \mathbf{k}_2)} \sum_{\mathbf{G}} \frac{4\pi}{[\mathbf{G} + T(\mathbf{k}_1 - \mathbf{k}_3)]^2} \\
& \times \langle i\mathbf{k}_1 | e^{i[\mathbf{G} + T(\mathbf{k}_1 - \mathbf{k}_3)]\hat{\mathbf{r}}} | a\mathbf{k}_3 \rangle_{\Omega_0} \langle j\mathbf{k}_2 | e^{-i[\mathbf{G} + T(\mathbf{k}_4 - \mathbf{k}_2)]\hat{\mathbf{r}}} | b\mathbf{k}_4 \rangle_{\Omega_0} .
\end{aligned}$$

In the first step the translated orbitals $\varphi_{a\mathbf{k}}(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\mathbf{R}}\varphi_{a\mathbf{k}}(\mathbf{r})$ reveal phase factors which reduce to Kronecker deltas when summing over all lattice vectors, see Eq. (A3). The space integrals over Ω are thus reduced to integrals only over the unit cell Ω_0 . The Kronecker deltas eliminate the \mathbf{k} sum over the BZ such that the Coulomb kernel leaves a sum only over the reciprocal lattice vectors. In the last step definition (26) was adopted.

Appendix B: Pseudocode for the parallelized implementation

Pseudocode of the parallel Γ -only implementation can be found in Fig. 8. The number \mathcal{N} of CPUs is divided into \mathcal{N}_B groups parallelizing over bands and \mathcal{N}_G groups parallelizing over plane-waves such that $\mathcal{N} = \mathcal{N}_B \mathcal{N}_G$. The set \mathcal{G} of all plane-waves is evenly divided into \mathcal{N}_G disjoint subsets: $\mathcal{G} = \bigcup_n \mathcal{G}_n$, $n \in [1, \mathcal{N}_G]$. Likewise the set of all occupied indices \mathcal{I} and virtual indices \mathcal{A} is evenly divided into \mathcal{N}_B disjoint subsets: $\mathcal{I} = \bigcup_n \mathcal{I}_n$, $\mathcal{A} = \bigcup_n \mathcal{A}_n$, $n \in [1, \mathcal{N}_B]$. Moreover, each subset of virtual band indices \mathcal{A}_n is further divided into \mathcal{N}_G disjoint subsets: $\mathcal{A}_n = \bigcup_m \mathcal{A}_n^{(m)}$, $m \in [1, \mathcal{N}_G]$, since for the calculation of the overlap densities the plane-wave parallelization is inapplicable. See Fig. 9 for an illustration.

```

Distribute occ. orbitals over  $\mathcal{I}$  among  $\mathcal{N}_B$  groups
Distribute virt. orbitals over  $\mathcal{A}$  among  $\mathcal{N}_B$  groups

# overlap densities (i)
for all  $i \in \mathcal{I}$  do
  fetch  $\varphi_i$  from other band groups
  for all  $a \in \mathcal{A}_{\text{this}}^{(i)}$  do
     $\rho_{ai}(\mathbf{r}) \leftarrow \varphi_a^*(\mathbf{r})\varphi_i(\mathbf{r})$ 
     $\rho_{ai} \leftarrow \text{FFT}[\rho_{ai}]$ 
    gather  $\tilde{\rho}_{ai}$  of this band group
    store  $\tilde{\rho}_{ai}(\mathbf{G})$  for  $\mathbf{G} \in \mathcal{G}_{\text{this}}$ 
  end for
end for

Distribute virt. orbitals over  $\mathcal{G}$  among  $\mathcal{N}_G$  groups
for all  $\tau$  do
  for all  $\mathbf{G} \in \mathcal{G}_{\text{this}}$  do
    # transformed states (iii)
    for all  $i \in \mathcal{I}$  do
      fetch  $\varphi_i$  from other band groups
      for all  $a \in \mathcal{A}_{\text{this}}$  do
         $C_{ia} \leftarrow e^{(\varepsilon_i - \varepsilon_a)\tau} \tilde{\rho}_{ai}(\mathbf{G})$ 
      end for
      gather  $C_{ia}$  of this plane-wave group
       $\tilde{w}_i(\mathbf{G}') \leftarrow \sum_a C_{ia} \tilde{\varphi}_a(\mathbf{G}')$  for  $\mathbf{G}' \in \mathcal{G}_{\text{this}}$ 
      redistribute  $\tilde{w}_i$  over  $\mathcal{I}$  among  $\mathcal{N}_B$  groups
      store  $\tilde{w}_i$ 
    end for

    # direct term (iv.a)
     $w_i \leftarrow \text{FFT}^{-1}[\tilde{w}_i]$  for  $i \in \mathcal{I}_{\text{this}}$ 
     $\varrho(\mathbf{r}) \leftarrow \sum_{i \in \mathcal{I}_{\text{this}}} \varphi_i^*(\mathbf{r})w_i(\mathbf{r})$ 
     $\tilde{\varrho} \leftarrow \text{FFT}[\varrho]$ 
    add  $\tilde{\varrho}$  of this plane-wave group together
     $e_d^{(2)} \leftarrow e_d^{(2)} + \sum_{\mathbf{G}' \in \mathcal{G}} \frac{4\pi}{G'^2} |\tilde{\varrho}(\mathbf{G}')|^2$ 

    # exchange term (iv.b)
     $w_i \leftarrow \text{FFT}^{-1}[\tilde{w}_i]$  for  $i \in \mathcal{I}$ 
    for all  $i \in \mathcal{I}$  do
      fetch  $\varphi_i$  from other band groups
      for all  $j \in \mathcal{I}_{\text{this}}, j \leq i$  do
         $\varrho_1(\mathbf{r}) \leftarrow \varphi_i^*(\mathbf{r})w_j(\mathbf{r})$ 
         $\varrho_2(\mathbf{r}) \leftarrow \varphi_j^*(\mathbf{r})w_i(\mathbf{r})$ 
         $\tilde{\varrho}_1 \leftarrow \text{FFT}[\varrho_1]$ 
         $\tilde{\varrho}_2 \leftarrow \text{FFT}[\varrho_2]$ 
        add  $\tilde{\varrho}_1, \tilde{\varrho}_2$  of this plane-wave group together
         $e_x^{(2)} \leftarrow e_x^{(2)} + \sum_{\mathbf{G}' \in \mathcal{G}} \frac{4\pi}{G'^2} \tilde{\varrho}_1^*(\mathbf{G}')\tilde{\varrho}_2(\mathbf{G}')$ 
      end for
    end for

     $E_d^{(2)} \leftarrow E_d^{(2)} + \frac{4\pi}{G^2} e_d^{(2)}$ 
     $E_x^{(2)} \leftarrow E_x^{(2)} + \frac{4\pi}{G^2} e_x^{(2)}$ 
  end for
  add  $E_d^{(2)}$  of this band group together
  add  $E_x^{(2)}$  of this band group together
end for

```

Figure 8. Pseudocode of the parallel Γ -only implementation. Hartree-Fock energies and orbitals for all occupied and virtual states are assumed.

- [1] C. Møller and M. S. Plesset, *Phys. Rev.* **46**, 618 (1934).
 [2] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (Dover Publications, Inc, 1996).

- [3] P. Y. Ayala and G. E. Scuseria, *J. Chem. Phys.* **110**, 3660 (1999).
 [4] M. Schutz, G. Hetzer, and H. J. Werner, *J. Chem. Phys.* **111**, 5691 (1999).

		bands			
		group 1		group 2	
plane-waves	group 1	\mathcal{G}_1	$\mathcal{I}_1 \mathcal{A}_1 \mathcal{A}_1^{(1)}$	\mathcal{G}_1	$\mathcal{I}_2 \mathcal{A}_2 \mathcal{A}_2^{(1)}$
	group 2	\mathcal{G}_2	$\mathcal{I}_1 \mathcal{A}_1 \mathcal{A}_1^{(2)}$	\mathcal{G}_2	$\mathcal{I}_2 \mathcal{A}_2 \mathcal{A}_2^{(2)}$
	group 3	\mathcal{G}_3	$\mathcal{I}_1 \mathcal{A}_1 \mathcal{A}_1^{(3)}$	\mathcal{G}_3	$\mathcal{I}_2 \mathcal{A}_2 \mathcal{A}_2^{(3)}$
	group 4	\mathcal{G}_4	$\mathcal{I}_1 \mathcal{A}_1 \mathcal{A}_1^{(4)}$	\mathcal{G}_4	$\mathcal{I}_2 \mathcal{A}_2 \mathcal{A}_2^{(4)}$

Figure 9. Illustration of the parallelization for the case of $\mathcal{N} = 8$ CPUs with $\mathcal{N}_B = 2$ band groups and $\mathcal{N}_G = 4$ plane-wave groups.

- [5] S. Saebø and P. Pulay, *J. Chem. Phys.* **115**, 3975 (2001).
- [6] B. Doser, D. S. Lambrecht, J. Kussmann, and C. Ochsenfeld, *J. Chem. Phys.* **130** (2009), 10.1063/1.3072903.
- [7] P. R. Nagy, G. Samu, and M. Kállay, *J. Chem. Theory Comput.*, acs.jctc.6b00732 (2016).
- [8] J. E. Subotnik, A. Sodt, and M. Head-Gordon, *J. Chem. Phys.* **128**, 034103 (2008).
- [9] F. R. Manby, D. Alf, and M. J. Gillan, *Phys. Chem. Chem. Phys.* **8**, 5178 (2006).
- [10] S. Casassa, M. Halo, and L. Maschio, *J. Phys. Conf. Ser.* **117**, 12007 (2008).
- [11] M. Halo, S. Casassa, L. Maschio, and C. Pisani, *Chem. Phys. Lett.* **467**, 294 (2009).
- [12] M. Halo, S. Casassa, L. Maschio, and C. Pisani, *Phys. Chem. Chem. Phys.* **11**, 586 (2009).
- [13] A. Erba, S. Casassa, R. Dovesi, L. Maschio, and C. Pisani, *J. Chem. Phys.* **130**, 0 (2009).
- [14] A. Erba, L. Maschio, C. Pisani, and S. Casassa, *Phys. Rev. B* **84**, 012101 (2011).
- [15] S. Casassa and R. Demichelis, *J. Phys. Chem. C* **116**, 13313 (2012).
- [16] E. Fabiano, M. Piacenza, S. D'Agostino, and F. Della Sala, *J. Chem. Phys.* **131** (2009), 10.1063/1.3271393.
- [17] L. Maschio, D. Usvyat, and B. Civalleri, *CrystEngComm* **12**, 2429 (2010).
- [18] P. Schwerdtfeger, B. Assadollahzadeh, and A. Hermann, *Phys. Rev. B* **82**, 205111 (2010).
- [19] K. D. Nanda and G. J. O. Beran, *J. Chem. Phys.* **137** (2012), 10.1063/1.4764063.
- [20] D. Stodt and C. Hättig, *J. Chem. Phys.* **137**, 114705 (2012).
- [21] F. Göltl, A. Grüneis, T. Bučko, and J. Hafner, *J. Chem. Phys.* **137** (2012), 10.1063/1.4750979.
- [22] M. Del Ben, M. Schönherr, J. Hutter, and J. VandeVondele, *J. Phys. Chem. Lett.* **4**, 3753 (2013).
- [23] M. Del Ben, J. VandeVondele, and B. Slater, *J. Phys. Chem. Lett.* **5**, 4122 (2014).
- [24] M. Del Ben, J. Hutter, and J. VandeVondele, *J. Chem. Phys.* **143** (2015), 10.1063/1.4927325.
- [25] S. Torabi, L. Hammerschmidt, E. Voloshina, and B. Paulus, *Int. J. Quantum Chem.* **114**, 943 (2014).
- [26] L. Hammerschmidt, L. Maschio, C. Müller, and B. Paulus, *J. Chem. Theory Comput.* **11**, 252 (2015).
- [27] Z. Kaawar, C. Müller, and B. Paulus, *Surf. Sci.* (2016), 10.1016/j.susc.2016.06.021.
- [28] S. Suhai, *Phys. Rev. B* **27**, 3506 (1983), arXiv:1011.1669.
- [29] J. Sun and R. Bartlett, *J. Chem. Phys.* **104**, 8553 (1996).
- [30] C. Pisani, M. Busso, G. Capecchi, S. Casassa, R. Dovesi, L. Maschio, C. Zicovich-Wilson, and M. Schütz, *J. Chem. Phys.* **122** (2005), 10.1063/1.1857479.
- [31] G. Kresse and J. Hafner, *Phys. Rev. B* **47**, 558 (1993), arXiv:0927-0256(96)00008 [10.1016].
- [32] G. Kresse and D. Joubert, *Phys. Rev. B* **59**, 1758 (1999).
- [33] M. Marsman, A. Grüneis, J. Paier, and G. Kresse, *J. Chem. Phys.* **130**, 1 (2009).
- [34] A. Grüneis, M. Marsman, and G. Kresse, *J. Chem. Phys.* **133**, 74107 (2010).
- [35] M. Katouda and S. Nagase, *J. Chem. Phys.* **133** (2010), 10.1063/1.3503153.
- [36] A. F. Izmaylov and G. E. Scuseria, *Phys. Chem. Chem. Phys.* **10**, 3421 (2008).
- [37] S. Y. Willow, K. S. Kim, and S. Hirata, *J. Chem. Phys.* **137** (2012), 10.1063/1.4768697.
- [38] S. Y. Willow, K. S. Kim, and S. Hirata, *Phys. Rev. B - Condens. Matter Mater. Phys.* **90**, 1 (2014).
- [39] M. Del Ben, J. Hutter, and J. VandeVondele, *J. Chem. Theory Comput.* **8**, 4177 (2012).
- [40] J. Almlöf, *Chem. Phys. Lett.* **181**, 319 (1991).
- [41] M. Häser and J. Almlöf, *J. Chem. Phys.* **96**, 489 (1992).
- [42] M. Kaltak, J. Klimeš, and G. Kresse, *J. Chem. Theory Comput.* **10**, 2498 (2014).
- [43] P. E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994), arXiv:arXiv:1408.4701v2.
- [44] J. Harl and G. Kresse, *Phys. Rev. B - Condens. Matter Mater. Phys.* **77**, 1 (2008).
- [45] J. J. Shepherd, A. Grüneis, G. H. Booth, G. Kresse, and A. Alavi, *Phys. Rev. B - Condens. Matter Mater. Phys.* **86**, 1 (2012), arXiv:1202.4990.
- [46] J. Klimeš, D. R. Bowler, and A. Michaelides, *J. Phys. Condens. Matter* **22**, 022201 (2010), arXiv:0910.0438.
- [47] D. L. Freeman, *Phys. Rev. B* **15**, 5512 (1977).
- [48] A. Grüneis, M. Marsman, J. Harl, L. Schimka, and G. Kresse, *J. Chem. Phys.* **131**, 2 (2009).
- [49] B. Ramberger, T. Schäfer, and G. Kresse, (2016), arXiv:1611.00689.